



Disrupting the dissertation: linked data, enhanced publication and algorithmic culture

Journal:	<i>E-Learning and Digital Media</i>
Manuscript ID	LDM-17-0015
Manuscript Type:	Learning in the Age of Algorithmic Cultures
Keywords:	Higher Education, Archaeology, Student Researchers, Linked Data, Semantic Web Technologie, Algorithms
Abstract:	<p>This article explores how the three aspects of Striphas' notion of algorithmic culture (information, crowds and algorithms) might influence and potentially disrupt established educational practices. We draw on our experience of introducing semantic web and linked data technologies into higher education settings, focussing on extended student writing activities such as dissertations and projects, and drawing in particular on our experiences related to undergraduate archaeology dissertations. The potential for linked data to be incorporated into electronic texts, including academic publications, has already been described, but these accounts have highlighted opportunities to enhance research integrity and interactivity, rather than considering their potential creatively to disrupt existing academic practices. We discuss how the changing relationships between subject content and practices, teachers, learners and wider publics both in this particular algorithmic culture, and more generally, offer new opportunities; but also how the unpredictability of crowds, the variable nature and quality of data, and the often hidden power of algorithms, introduce new pedagogical challenges and opportunities.</p>

SCHOLARONE™
Manuscripts

Disrupting the dissertation: linked data, enhanced publication and algorithmic culture

Introduction

This article describes in detail a specific pedagogical setting, namely project work in undergraduate archaeology, and analyses this through the lens of Striphas’ notion of ‘algorithmic culture’ (Striphas, 2009, 2015). Specifically, it draws on the experience of developing and evaluating the educational potential of semantic web technologies to enhance pedagogical practice in higher education, as part of a large funded research project. ‘Ensemble: Semantic Technologies for the Enhancement of Case Based Learning’ (Carmichael *et al.* 2012; Martinez-Garcia *et al.* 2012). It describes how the idea of ‘algorithmic culture’ has in education been hitherto explored, before discussing how the semantic web, both as a broad vision and as a set of technologies and approaches, aligns with this. It then presents the setting and the ways in which the potential of semantic web technologies was investigated, with an emphasis on their implications for student research and the writing associated with it. The article concludes with a discussion of the pedagogical implications more generally, as well as some reflections on how thinking and learning within algorithmic cultures might serve as a useful starting point for critical studies of education and a framework for the development of learning technologies.

Algorithmic culture in education

The term ‘algorithmic culture’ has come to prominence recently due to the work of Striphas (2009, 2015) who describes how combinations of digital technologies “engage in what’s often considered to be the traditional work of culture: the sorting, classifying, and hierarchizing of people, places, objects, and ideas” and then “feeds back to produce new habits of thought, conduct, and expression that likely wouldn’t exist in its absence - a culture of algorithms” (cited in Granieri, 2014). Drawing

on Williams (1983) notion of ‘keywords’, whose changing meanings signal underlying epistemological and ideological currents, Striphas argues not only that culture has been transformed by digital technologies but that three key ideas - ‘information’, ‘crowds’ and ‘algorithms’ - have also been imbued with new connotations (Striphas, 2015: 395). Striphas also highlights two contradictions: first, between the rhetorics of openness, transparency and ‘global public goods’ that have accompanied the growth of digital networks, and the reality of privatization, ‘paywalls’ and closed systems; and second, between the ‘crowd’ with its connotations of a participatory and democratic public culture which, he suggests, is merely a placeholder for the ‘algorithm’ which is the most dominant force shaping contemporary culture (Striphas, 2015: 407). Kitchin (2016) and Neyland and Moellers (2016) discuss multiple theoretical perspectives on the social role and power of algorithms, highlighting, as Striphas does, contradictions between rhetorics of openness and participation, and the power relations that algorithms establish, reinforce and maintain.

Debates about the influence of algorithms and ‘code’ have been ongoing in education for some time: Peters (2011) argues that education is a “core site” for an emerging algorithmic and cognitive capitalism which is centralising and monetising the internet and seeks to exclude, marginalise or monetise information sources while channelling, monitoring and commoditizing social interactions. Peters (2011) and Dyer-Witthford (2005) both draw on autonomist Marxist theory which takes as its starting point Marx’s argument in the *Fragment on Machines* (1973: 692-711) that alienation from work is accelerated and new social relations formed through automation. This in turn leads to the isolation and alienation of members and subgroups of what Striphas would term the ‘crowd’ (see, for example, Berardi, (2005, 2009), for more extended discussion of these processes and their outcomes). A similar argument is advanced by Selwyn, who suggests that education is increasingly treated as a computational project, characterized by algorithmically driven systems thinking whereby “complex (and unsolvable) social problems associated with education can be seen as complex (but solvable) statistical problems” (2015: 72). In the same way, claims about the benefits of large-scale

data collection and aggregation as a means to enable the harmonisation and standardisation of education systems (Ozga, 2012) and developments such as learning analytics in online environments, adaptive assessment systems, target setting and learning outcome tracking, have been critiqued as being associated with top-down managerialist approaches (Ozga, 2009; Selwyn 2015).

In contrast with these broader studies of education systems, and accounts of how ‘personalisation’ of educational experience might be enabled, there is little research on how algorithmic culture impacts on pedagogical practice or curriculum design. Williamson (2015a, 2015b) discusses the role of algorithms and code in the digital governance of education, and highlights the challenge and contradictions of “tracking and monitoring students as data objects [while] enabling students to become active citizens who can participate in the practices and performance of programmable urban processes” (Williamson (2015c: 2). This concern to educate what might be termed ‘citizen data scientists’ (following Irwin’s (1995) vision of ‘citizen scientists’ able to participate in scientific activities and engage in science policy processes) is, in the UK, evidenced in the recent National Curriculum in Computing (Department for Education, 2013) and associated initiatives which highlight the importance of ‘learning to code’ (Williamson 2016). Edwards and Carmichael (2012) point to the role that coding plays in curriculum content and potentially in establishing and reinforcing a ‘hidden curriculum’ of highly specific, targeted content, where the assumptions and intentions of curriculum developers may be concealed from both teachers and learners.

There have, of course, been studies of technologies which describe the *manifestations* of the elements of Striplas’ algorithmic culture. Examples include studies of social software in education (see for example, Bennett *et al*’s (2012) study of the ‘practice logics’ of social software and higher education); of collaborative authoring in wikis (Vratulis and Dobson, 2008), crowd-sourcing of educational content (Corneli & Mikroyannidis, 2012; Porcello & Hsi, 2013), and the concerns that teachers and students have about information overload (Bawden and Robinson, 2008). However,

there are few studies that explore, either empirically or in theorised accounts, how algorithmic tendencies in particular enable, direct, or constrain pedagogical practices and dialogues. This may reflect the very issue that Striphas highlights: that the dominant discourse in education emphasises ‘the crowd’ amid rhetorics of openness, participation and production, while paying insufficient attention to the role of algorithms (Striphas, 2015: 403).

The Semantic Web and Algorithmic Culture

The Semantic Web or ‘Web 3.0’ is a re-envisioning of the World Wide Web into a global library that “not only links documents to each other but also recognises the meaning of the information in those documents” (Frauenfelder, 2001). It incorporates all the elements that Striphas identifies, in that it seeks to make available diverse, distributed, networked information to a global community of users who may be consumers, contributors or both, and in that standardisation of the representation of that information should make it possible for it to be searched, selected, aggregated, ‘reasoned across’ and presented by software. Specific semantic technologies and approaches such as metadata schemes, ontologies and taxonomies, semantic search tools, recommender systems, and machine based reasoning are the elements that drive algorithmic culture in education and beyond. So the semantic web as a broad vision can be seen as a particular manifestation of algorithmic culture, while semantic technologies are located, often invisibly, at the heart of that culture as it is experienced by users of internet services, including educational applications and platforms. Such applications include ‘recommender systems’ which suggest resources and products that might be of interest to users; in higher education, examples include reading list management software such as *Talis Aspire* (<http://talis.com/reading-lists/>).

Before looking in detail at our particular example, three themes are worth exploring regarding the alignment of the semantic web and algorithmic culture. The first is related to the nature of the semantic web as a ‘grand challenge’ in computing sciences and the way that this has evolved; the

second relates to the central role in semantic web developments of ontologies and taxonomies, and their relation to ‘crowd-sourced’ knowledge; and the third concerns the way in which the semantic web and its associated technologies have been understood and implemented in educational and research settings.

Early accounts of the semantic web stress the potential of relatively autonomous algorithms or “software agents” to traverse diverse, distributed data in order to offer end users apparently seamless and personalised services. Berners-Lee, Hendler and Lassila (2001) describe such scenarios related to personal healthcare management, as do Anderson and Whitelock (2004), Koper (2004) and Lytras and Naeve (2006) in relation to potential educational implications of this broad vision of a future semantic web. Subsequently, the emphasis has shifted towards the potential for specific semantic web technologies being integrated into other software environments, platforms and search tools, and the grand vision of a next-generation web has been eclipsed both by the rise of social software (although many of these incorporate semantic web technologies) and by a more limited ‘Linked Web of Data’ (Bizer, Heath and Berners-Lee, 2009; Heath and Bizer, 2011) that draws on some of the more robust and widely-adopted semantic web technologies and which has also been encouraged by ‘open data’ initiatives such as those adopted by government agencies in the US (<http://www.data.gov>); the UK (<http://data.gov.uk>) and in supra-national networks such as the European Research Area.

A significant area of semantic web development has been the development of metadata schemes, technical ontologies and taxonomies to enable consistent description of online content, and, therefore, to allow searching, retrieval, aggregation, interoperability and machine-based reasoning across distributed digital networks. This leads, however, to a tendency to see the information so described as neutral or uncontested, with its associated metadata informing rational choices (by machines or people) and widespread communication. Ontologies and taxonomies are by their nature

reductive as they express consensus understandings and are typically reached through consultation with an expert community (as described by Ribes and Bowker (2008) in their account of the development of a shared ontology for the geosciences) at the expense of detailed contextual or cultural nuances. In the same way, taxonomies used within the medical sciences such as the World Health Organisation *International Classification of Diseases (ICD)* (<http://www.who.int/classifications/icd/>) and *International Classification of Functioning, Disability and Health (ICF)* (<http://www.who.int/classifications/icf/>) may be useful for the identification of individuals with particular conditions or disabilities, conducting systematic literature reviews or meta-analyses, or identifying demographic patterns, but they are not sufficient to represent the particularities and subjectivities of their circumstances or experiences. That said, there are many examples of local ontologies, taxonomies, keyword thesauri and ‘folksonomies’ that do reflect local, situated practice and knowledge representations, with these articulating with more formal ontologies and taxonomies, and sometimes (but not always) becoming incorporated into them, often in the context of a hybrid social semantic web (Mika, 2007a; 2007b). Features of such expert taxonomies and ontologies, or hybrid social semantic systems which combine patterns of user activity with those more formal representations of knowledge means that the semantic web as a vision, and systems that incorporate even just some semantic web technologies, could be seen as implementations of, or even as driving the development of, algorithmic culture. What is particularly significant in relation to Striphas’ concern about the relative power of algorithms and crowds is the fact that the role of expert ontologies, machine reasoning and algorithms may (like the intentions of the curriculum developers mentioned above) be concealed in online environments that emphasise social activity, participation, and user contributions.

The final respect in which the idea of algorithmic culture aligns with or echoes patterns in the development of the semantic web relates to the nature of the models and exemplars that are offered in relation to educational settings and practice. In the course of the *Ensemble* project, it often proved

difficult to engage teachers and learners with the complexities of the emerging semantic web and a number of strategies were developed to make key concepts more tractable and affordances (and challenges) more evident. These included a range of participatory design approaches, rapid prototyping, and envisioning activities based on pedagogical requirements rather than technological opportunities. The difficulties encountered were compounded by a lack of accessible examples or use-cases related to pedagogical practice, with the majority being concerned with educational administration and management, resource management and retrieval, or assessment and learner profiling. Carmichael and Jordan (2012), in the introduction to a special issue of *Technology, Pedagogy and Education* in which a number of papers from the project were presented, argued that for semantic web technologies to be more widely adopted in educational settings, greater emphasis needed to be placed on the exploration of current pedagogical practice and how this might be enabled or enhanced by selective application of semantic web technologies. Moving forward to the current interest in algorithmic culture and computational thinking, we see similarities with this prior experience in that, while there are broad accounts of how these might affect educational systems, detailed accounts of the implications for educational practice, rather than governance or management, are limited in number. This article, then, draws on the extensive empirical record of the *Ensemble* Project, in order to explore how a specific instance of an emerging algorithmic culture, based on a set of semantic technologies, might impact on teaching, learning and assessment practice, particularly the writing of dissertations, in one particular educational setting - undergraduate archaeology.

The ‘Artefact Project’ as Dissertation and Research Object

Final year undergraduate students are often required to produce a research dissertation based on a literature review or a practice-based project that demonstrates both subject knowledge and a mastery of skills. For example, the European Qualifications Framework requires that achievement at this

level involves “advanced knowledge of a field of work or study, involving a critical understanding of theories and principles” as well as “advanced skills, demonstrating mastery and innovation, required to solve complex and unpredictable problems” (European Commission, 2016). Healey et al. (2012) suggest that dissertations should reflect both changing research practices and a globalised workplace, and they argue that while online technologies have been key influences in terms of changing other aspects of higher education practice, there has been only limited development in relation to dissertations. These typically continue to reflect established forms of writing within disciplines, and so are often based on the format and structure of journal articles, evaluation reports, business case studies or service improvement proposals, and while they may be encouraged to demonstrate specific ‘techno-literacies’: conducting online searches, analysing data, correctly citing source materials, practices related to data use or reuse either in academic research or professional practice are less likely to be used.

In the course of the *Ensemble* project’s work, undergraduate archaeology at one of the universities involved in the project had emerged as an educational setting in which there was a potential role for semantic web technologies in support of some form of case based learning. Initially, work focussed on the role of search and visualisation tools in the context of fieldwork, the site being the ‘case’ (this was a strand of project development that was later developed with teachers and students of environmental sciences; other settings in which the project undertook research and development work included marine operations and management, plant sciences, contemporary dance and education studies: see Martinez-Garcia et al. 2012 for details).

Subsequently, the dissertation project undertaken by final year undergraduates became a focus for development. This was known as the ‘Artefact Project’ as some students did write detailed accounts of single artefacts, but others focussed on sites, extant buildings, assemblages such as hoards, or the distribution of artefacts such as coins. As a dissertation, the project required that students

demonstrate subject knowledge but also that they presented an account similar to that which might be published in an academic journal, or as a publication produced by a museum, county record office or archaeological society. It was the geospatial elements of such projects that at least initially, appeared to be a good match for some of the semantic technologies with which the project team were working. In particular, the *Exhibit* lightweight semantic web publishing framework (Huynh, D., Karger, D., and Miller, R., 2007) was being used effectively to develop prototype semantic web applications in which diverse data were aggregated and presented for classroom use through search interfaces, image libraries, timelines, and interactive maps. As part of a detailed account of the working practices of the project team, Rimpiläinen (2012: 223-229) describes how the idea of a ‘Data Aggregating Document’ or ‘DAD’ (in fact, a content-rich, highly interactive web page) emerged as a means of preserving the narrative structure of Artefact Project dissertations while enhancing them through the use of data ‘inclusions’, visualisation tools, and other interactive content.

One of the stimuli for this shift away from producing discrete web applications (such as maps or timelines) to incorporating semantic technologies into documents was the work of Shotton et al. (2009) who present an enhancement of an already-published journal article in the field of biomedical science using semantic web approaches. They argue that research articles serve as a “conduit to the publication of research data” (2009:1), although this may reflect a particular disciplinary perspective that emphasises empirical data over narrative interpretation or argumentation. Other authors present a rather broader range of approaches to enhanced publication, stressing the interlinkage of elements of publications and the opportunities this offers for reanalysis and post-publication discourse (Hoogerwerf et al., 2013; Jankowski et al., 2013). However, De Waard (2005) highlights resistance to change in academic publishing in the face of the increasingly digital nature of scientific work, particularly where preferred narrative structures make it difficult to deconstruct texts into modular elements for subsequent reconstruction into knowledge networks of the kind enabled by semantic

1
2
3 web technologies. In the course of the *Ensemble* project this resistance was encountered on
4
5 occasion: while teachers in some disciplines saw the potential for student assignments to be reframed
6
7 as technical reports or presentations (incorporating knowledge objects such as linked data sets,
8
9 visualisations or dynamic content), in others the enduring dominance of the argumentative academic
10
11 essay as the focus of assessment made this less appealing or less easy to implement in practice. This
12
13 was less a reflection of specific technical concerns, but rather of the fact that student dissertations
14
15 represented the culmination of a process of training in academic writing that would have to be
16
17 comprehensively revised.
18
19

20
21 There has also been a recognition that new forms of representation and publication may be required
22
23 to reflect the growth of a new paradigm of scientific research, ‘eScience’, where scientists use large
24
25 scale complex instruments to relay large volumes of data to data centres for processing by software
26
27 and analysis by distributed teams (Gray, cited in Hey, Tansley and Toll, 2009: xviii). Fox and
28
29 Hendler argue that such eScience, involving very large-scale data, requires ‘semantic based
30
31 methodologies, tools and middleware’ to enable data analysis and knowledge modelling to be
32
33 accessible to scientists, students and non-experts (Fox and Hendler, 2009: 148), with frameworks
34
35 such as *Exhibit* being examples of such tools. De Roure suggests that the traditional format of most
36
37 research publications is at odds with such “data-driven or data-intensive science” where humans,
38
39 specialist equipment, and the data that they generate, are widely distributed across networks and
40
41 publication may involve dispersed teams in analysis of data from multiple sources (2014: 234):
42
43 obvious examples include climate change and epidemiological research. They propose that a better
44
45 fit for research dissemination in this new paradigm is that of the aggregated and shareable “research
46
47 object” (2014: 236) of which the *Ensemble* data-aggregating documents can be seen as examples.
48
49 Such objects are ‘semantically rich’ and layered on top of information provided as linked data
50
51 (Bechhofer et al., 2013). Appropriately constructed, they can present the richness and complexity of
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

research data alongside the description and discussion of theory and conclusions, and enable reanalysis and collaboration.

In these developments, we see Striphas’ idea of algorithmic culture as it manifests in the context of ‘big data’ and ‘eScience’ (Striphas 2015: 396), with information, researchers and research users, and algorithms, all distributed across networks, with enhanced publications or research objects representing points of intersection and intensity around which academic discourse is focussed. What the *Ensemble* project’s work in undergraduate archaeology allowed was the exploration of these themes in the context of much smaller scale student research activities designed to produce research objects which also had pedagogical purpose, and which contributed to the assessment of those students.

Following the broad approach used by Shotton et al. (2009) in their reconstruction of an already-published journal article, several existing Artefact Project dissertations were reviewed to assess how they might be reconstructed as interactive data-aggregating documents. These included studies of single buildings where students had manually drawn maps and illustrations to show the history of their construction, habitation, and of collections of items such as mediaeval coins which had been minted in different locations but which had then been distributed in a wide geographical area through travel and trade. Projects such as these would have lent themselves to representation through interactive timelines or maps respectively, but the example that was explored in most detail involved a study of a small collection of Anglo-Saxon brooches that offered opportunities to present information using a range of presentation and visualisation tools. This study was developed first as a technical demonstrator, and was then used as the focus of interviews with teachers and curators involved in supporting students undertaking Artefact Projects, and it is this example that forms the focal case for the remainder of this article. Rimpiläinen (2012: 230-243) again provides a very detailed account of the technical processes by which existing paper-based dissertations were

translated into data-aggregating documents. Here, we draw attention to the processes of translation and provide an example of the potential of semantic technologies in this particular pedagogical setting.

Reconstructing the Artefact Project: The Girton Anglo-Saxon Brooches

The original Artefact Project was submitted for assessment as a 24 page word-processed report which described three Anglo-Saxon brooches that had been discovered in a grave uncovered, along with other remains, during the building of Girton College, Cambridge, in 1881. It describes the brooches and includes hand-drawn illustrations of them, along with a discussion, based on secondary sources, of how they might have been worn and by whom, and of their cultural significance. It also describes the site and the history of excavations there, including photographs, sketches and maps, and, using a map, locates it in relation to broader patterns in the distribution of similar finds and evidence of settlement and cemetery distribution across the East of England during the Anglo-Saxon period.

Drawing on the *Exhibit* web framework, the elements of the original document were translated into elements of an interactive web page. Narrative sections were generally left as static text, though when the history of excavation at the site was discussed, these were represented as interactive timelines presenting the same information as the original narrative. Other sections of the document were translated as described in Table 1 with some of the potential further developments that were discussed, but were not implemented, also identified.

[TABLE 1 about here]

A section of the data aggregating document that includes some minimally altered text along with the timeline of excavations is shown in Figure 1.

[FIGURE 1 about here]

The map in Figure 2 is shown having already been manipulated using the faceted search tools, with a site selected, its details and a link to an external catalogue which contains a full record appearing as a ‘pop-up’ box. What is most significant in this interactive element is that, in addition to sites identified in the original project document, records of all of the Mediaeval finds and sites (including Anglo-Saxon examples) in the university museum database have now been included, together with a faceted search tool that allows these to be selectively displayed according to period and type.

[FIGURE 2 about here]

As a demonstrator of the potential of semantic web technologies and linked data approaches, the translated document, like the example presented by Shotton et al. (2009), proved useful. It not only showed what appropriately structured information, standardised metadata, and web development frameworks could enable; it also highlighted obstacles such as ‘legacy databases’ being inaccessible, inconsistencies in formats and descriptive metadata; and the problems caused when data from disparate sources, collected and described according to different protocols, were aggregated into larger databases or repositories. But to interpret these, and to seek to address them, purely as technical issues, would fail to address the practices of research, archiving, curation, and sharing, as well as those of the researcher or writer seeking to reuse extant data in new productions such as an enhanced Artefact Project, dissertation or research article. In Striphas’ terms, this would risk privileging information at the expense of human participants and their practice, and overlooking some of the hidden influences of algorithms entirely, and we were to discover this in the course of a series of interviews with curators, teachers and students about the project and the implications of reframing it as the production of an enhanced semantic hypertext.

Some Pedagogical Implications of the Data-Aggregating Document

By the time interviews took place, members of the project were used to presenting prototypes, some of which implemented sophisticated semantic web technologies, to teachers and students across a

range of disciplines, only to be told that this was not what they wanted at all, and new and often highly divergent ideas were proposed. As a result, participatory design and development practices extended beyond conventional requirements gathering and involved rapid development of prototypes which were then used as the focus for further discussions (Carmichael and Litherland, 2012; Tracy and Jordan, 2012). It was in this context that these interviews took place. At the same time, our expectations were largely framed in terms of how semantic web technologies, and in particular the ability to incorporate linked data might contribute to the development of student projects as ‘enhanced publications’.

What emerged from these interviews were several themes that provided insights into how semantic web technologies might be implemented in the context of the Artefact Projects, but also, aspects of a broader disciplinary algorithmic culture. The first of these concerned data quality and accuracy.

When the map shown in Figure 2 was inspected, it was evident that there were locations where large numbers of objects appeared to have been found; however, inspection of the full museum catalogue entries suggested that these were not hoards, but rather a product of dispersed locations being recorded initially as a town or village name, and latterly been converted to a single geospatial ‘data point’ - usually close to the town or village centre. Further, initially mystifying, examples suggested Anglo-Saxon finds off the coasts of West Africa and close to the Seychelles. These were not an indication of a much wider Anglo-Saxon sphere of influence than is conventionally believed to be the case: rather, in the first case, they were a product of ‘null’ values being converted (at some point) into zeros and then being interpreted by the mapping algorithm as “0 degrees North, 0 degrees East”; and in the second, a transposition of the values in the geospatial data for a site near Cambridge (around 52 deg. N, 0 deg. E) giving the rather more tropical location of (0 deg. N, 52 deg. E).

Rather than simply seeing these as data errors, which were more easily observed in the map than in a catalogue of data entries, an interview with Michael, one of the museum curators, provided insight

into to the social practices that led to these patterns. While original paper records often included detailed information about the location at which the artefact was found, the curator responsible for the online database sometimes purposefully excluded specific details about the location, including its geospatial data, in order to avoid encouraging amateur historians and metal detectorists from visiting them in search of further valuable objects. He explained “... of course, online it doesn't tell you exactly where it is ... that's about security”. Regardless of the apparent ease with which data sharing, algorithmic transformation and online representation could be enabled, the format, detail and accessibility of the information in such cases reflected the perceptions of a powerful individual as to the potential uses and misuses of the data. In Michael’s words, “power and control defines what most museums do”. The version of the database that was available to the public, and as a data source for linked data applications, contained place names, which had then been converted into approximate geospatial data using an online tool. It is these locations that appear in Figure 2, and this partly manual, partly algorithmic process probably accounts both for much of the apparent clustering of archaeological finds close to town centres and other points of interest, and for the accidental transposition of latitudes and longitudes by a data entry administrator. The apparently simple process of information sharing is here complicated by the role of one algorithm (converting names to approximate geospatial data) in obfuscating and concealing information, while another, built into the Google Maps code within the Data Aggregating Document, is used in an attempt to make meaning from it.

The second theme to emerge in the course of the interviews was the importance of narrative both as a framing of enquiry and as a representation of its outcomes. Many Artefact Projects began not with the intention of conducting a large-scale survey but rather with tracing the history of a location, or the discovery, curation and interpretation of an object, or involved the student piecing together partial evidence from multiple sources. Sue, one of the museum curators who also acted as a tutor on the Artefact Project was concerned that automation of data collection might mean that “the whole

creative process loses out ... what's so good about doing these projects is finding some little snippet of information which is tucked away in a local collection or in a local library and drawing all of those things together." This concern with narrative was not simply attributable to a preferred mode of representation; rather, it was a necessary element of archaeological practice with which students needed to engage. As Sue described, this practice has evolved, partly as a result of new technologies being developed and adopted, " ... when people are doing their research, different aspects of the object become more or less important. So when 50 years ago you would say that is a great square headed brooch and that is all anybody needed to know, now we are wanting to know what is on the back of it or how many bits of enamel are on it or is it Salin's Style 2 [a categorisation of animal forms in Anglo Saxon art] ... then in the future people might be wanting to analyse the metal content ... but 20 years ago nobody was interested in that."

What this meant in practice was that any artefacts, and equally, the information and interpretations that accompanied them, had to be understood in the context of their discovery, description, curation and preservation. Despite the apparent ease with which metadata describing objects could be collected and aggregated, the taxonomies and other information that would be used by algorithms to present information to users of web based systems (including, of course, the data aggregating documents) might be based around multiple taxonomies and different notions of what was salient or noteworthy about the object, with curators being very influential in how collections were catalogued and objects described. Sue explained that "in [the] database where there clearly was a collector who covered a particular area ... [the artefacts] have been coded and classified, possibly even mistakenly classified, in some cases according to their particular perspective". In other cases, the work of individual archaeologists in developing taxonomies and typologies (including one for the classification of brooches, referenced in the Girton Artefact Project) had the potential to be accepted as the *de facto* standard - the type of information that might subsequently be coded into a database or algorithm.

As such, tools that employed algorithms to aggregate metadata from different sources or different periods, or which attempted to map diverse taxonomies into a set of ‘preferred terms’ were seen as being, if not unhelpful, something that should be treated with caution. Michael stated that “translation is always a form of transformation ... basically you lose everything you wanted in the first place ... everything you wanted in terms of diversity is gone”. He described how his concern was to try to reduce the levels of interpretation required for classification of objects, but he recognised that this was in opposition to the goals of many large scale multi-national projects which were seeking to establish agreed and stable meanings for terms in their taxonomies in order to enable reciprocal sharing of data and searches across multiple databases, often using semantic web tools.

Archaeology is perhaps a case in which the issue of who comprises Striplas’ ‘crowd’ and what constitutes the ‘information’ that is generated is particularly salient; as well as the wide range of ‘citizen scientists’ working as amateur archaeologists through local museums, for example, there exists an enthusiastic metal detectorist community with its own patterns of technology use and data sharing practices. These articulate with more formal academic activities through initiatives such as the Portable Antiquities Scheme (<https://finds.org.uk/>) which allows participants to publish records complete with detailed geospatial data. Such involvement goes beyond the scope of most ‘crowdsourcing’ initiatives, and further blurs the distinction between producers and users of information, even before considering what role algorithms might play: Would they provide a means of combining information from across this diverse crowd? Or maintaining security, as Michael suggested? Or is there potential for both, either in some kind of coexistence, or as rival cultures?

A final theme which emerged represented a pragmatic response to some of these complexities on the part of teachers and students who set pedagogical boundaries so they could engage with a sufficiently challenging topic that they could demonstrate knowledge and domain-specific skills while maintaining some control over the scope of the project, and teachers were cautious about their

undertaking large-scale or unwieldy projects. This concern emerged in other *Ensemble* project settings too: Edwards, Tracy and Jordan (2011) discuss the implications of teachers' reluctance to bring large scale, untrusted data into teaching activities.

One aspect of this control and boundary setting was manifest in students' concern to select topics that represented opportunities to demonstrate their competence and produce an engaging report, while remaining manageable in terms of the volume, complexity and accessibility of relevant data. Despite the emergence of linked data approaches, many archaeological databases are still not accessible online, and require visits to county record offices and other archives, so students were wary of attempting projects that were too broad in scope, and acknowledged that they carefully selected artefacts or localities based on the range and quality of data related to it. One described how "The Iron Age is quite complicated and there is quite a lot of information about it, so it is more about sifting down to see what is feasible, rather than struggling to collect data which is difficult to get at".

Information, Crowds, Algorithms ... and Dissertations

Much of the work of the *Ensemble* project was about envisioning new educational activities, opportunities and tools, and the projects' work in archaeology was no different. At the time that the project first engaged with the idea of dissertations as enhanced publications, what emerged was a superficially appealing argument which suggested that, in fields where engagement with open data, big data, or data visualisation is becoming the norm, or where academic practice is beginning to adopt models of enhanced publication of the kind we have described here, replacing the requirement to write a conventional dissertation with the production of a research object is a potential means bridging the gap between taught courses and employment or research.

However, seeing such emerging practices (whether in archaeology or other educational contexts) simply as a new set of techno-literacies related to the use of open and linked data, web-based

authoring environments and data visualisation tools, and reflecting these in revised course requirements and assessment criteria, risks falling into the trap that the authors did as they set out to ‘enhance’ the Artefact Project. As Bayne (2015) argues, the rhetoric of ‘enhancement’ often conceals a ‘constellation’ of social, technological and ideological interrelationships and encourages education and technologies to be seen in an instrumental and performative way.

Seeing the dissertation or research object as an object, node or point of intensity in a web of information stresses informational aspects at the expense of social practices and may miss opportunities to explore or interrogate algorithmic elements altogether. Such an approach leads to the kinds of issues that emerged through our work in archaeology being interpreted in terms of the credibility of sources (of information), or reliability or accuracy (of information), and would see tools such as taxonomies, ontologies and metadata schemes solely as neutral information exchange mechanisms, rather than recognising them as hegemonic, often contested and frequently frustrating of efforts to represent situated practice or subjectivities. And from this it follows that the assessment of any such enhanced dissertation would simply require students to demonstrate an ability to engage with data sources and metadata, assess their credibility and accuracy and use them to ‘enhance’ their publication as a standalone, summative and static production.

A more ambitious and disruptive approach would invite students to consider the implications for the production of their dissertation (which can still fulfil a bridging role) of being part of an informational web of data, but would additionally highlight the changing nature of that web. By encouraging students to recognise the influence of human actors and algorithms in generating, translating and making available new information relevant to their own dissertation or research object, they would be required to consider various contingencies and to see their work as tentative and conjectural. The map in the Girton Artefact Project, for example, would continue to develop as new objects are found, catalogued and their metadata published; more accurate geospatial data might

1
2
3 be made available; a metal detectorist might discover an Anglo-Saxon burial site with grave goods
4 similar to those found at the Girton site; or new data about the composition of alloys or the
5 provenance of semi-precious stones might cause reinterpretations of the artefact, site or broader
6 context. Students would need to establish what events, and what evidence, might challenge or
7 confirm their current interpretations, or generate new questions and lines of enquiry.
8
9

10
11
12
13
14
15 More challenging, then, for the student would be a consideration of how the various algorithms that
16 operate within this web, linking data sources, operationalising taxonomies and ontologies, and
17 displaying data might change, and what the implications of these changes might be. What if, for
18 example, an algorithm that currently obfuscates geospatial data so that only approximate locations
19 are published (the human partly responsible data conversion having been replaced) is supplanted by
20 one that publishes precise data for all finds, other than the most significant examples? Would points
21 on the map moving, by 500 metres perhaps, have bearing on their current interpretations and
22 hypotheses? Students' (and authors more generally) acceptance that they are not entirely in control
23 of information with which they are working will be key to working effectively in such
24 circumstances, and will also need to be reflected in assessment practices. Most critically, as Borg
25 and Boyd Davis (2012) argue in relation to doctoral work, research dissertations are contingent and
26 changeable, and technological developments mean that both their format and content are subject to
27 change. This is at odds with the expectation that dissertations need to be summative and static in
28 order to be reliably assessed.
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45

46
47 What a sensitivity to the idea of algorithmic cultures offers to educational scenario-building, is not
48 restricted to the opportunities or implications of data linkage, but rather directs attention to
49 alignments and tensions between social and technical aspects; the development and reproduction of
50 cultural practices, and the asymmetrical and often power-laden relationships between different
51 participants, including the student themselves. It was only once we began to explore the complex
52
53
54
55
56
57
58
59
60

and in some cases hitherto concealed practices, ‘hidden hands’ and algorithms at work that we began to gain insights into how semantic technologies (over which we had a degree of control) intermeshed with the broader algorithmic culture of archaeological research, curation and teaching over which we had no control whatsoever. This therefore, should also be the concern of the student, and needs to be reflected in the design of educational activities (dissertations included) and in their assessment.

It follows that, if they are to engage in disciplinary or professional communities, students will need to be aware of the influence of the ‘crowd’ (and its constituent parts) on the classifications, structures and algorithms that affect the information with which they work. Proposed explanations and analysis will need to include reflections on the power and influence that they themselves do or do not have over information and algorithms. The tentative analyses discussed above could be linked to other dissertations on which they build, which they critique, or which are related to the same topics, to create a web of explanations that show alternative perspectives and lines of enquiry. The culture that operates in the field of archaeology education is particularly influenced by the tentative and shifting nature of knowledge in a discipline where new discoveries can overturn established knowledge structures and interpretations, whilst others remain unchanged for years. These processes and stable elements themselves would become a legitimate, and even critical, focus of enquiry if students are to develop an appreciation of their own role within that culture.

Concluding Remarks

This brings us to a point where we can make a number of conclusions and suggestions. Firstly, that if some form of enhanced dissertation were to be introduced (in archaeology or in other settings), it would need to be accompanied not by support for the development of discrete techno-literacies, but rather by an expansive, critical form of information literacy. The research object, in other words, needs to be accompanied by a critical reflection on its production and on the ways in which social, cultural and technological practices impact on individual and collective, disciplinary or professional

practices. This would involve comprehensive reworking of curriculum aims, assignment briefs and assessment criteria to reflect a disposition toward critical enquiry and engagement with these different kinds of knowledge.

A more general point is the importance of recognising that, despite tendencies in this direction, algorithmic culture should not be able to operate invisibly, to be ‘black-boxed’ or to be subsumed by rhetorics of ‘crowdsourcing’ and ‘openness’. Students, whether engaging with, or engaging in, research, need to be supported to recognise the role of the human and non-human agents at work in educational and professional settings, and to explore the new “habits of thought, conduct and expression” that these enable. We have discussed these issues in the context of dissertation projects in one very distinctive educational context, but they are far more wide-reaching, as university courses evolve in order to reflect how emergent technologies are increasingly defining professional practice and shaping workplace environments.

Finally, given the concerns expressed about the adequacy of theorisations of the role of technology of education and specifically in pedagogical practice, the notion of algorithmic culture as articulated by Striphas (2009, 2015), with its emphasis on interrelations and tensions between social, informational and algorithmic seems promising. In the particular context of educational semantic web technologies, which have hitherto lacked compelling theoretical framings beyond those concerned with information management and personalisation of services, the debate initiated by Striphas and others (and to which we hope this article has contributed) also holds the promise of better theorization of these still-emergent technologies and the practices and discourses that accompany them.

References

Anderson T and Whitelock D (2004) The Educational Semantic Web: Visioning and Practicing the Future of Education. *Journal of Interactive Media in Education* 2004(1). Available at: <http://www-jime.open.ac.uk/articles/10.5334/2004-1/> (accessed 8 May 2017).

Bawden D and Robinson L (2008) The dark side of information: overload, anxiety and other paradoxes and pathologies *Journal of Information Science* 35(2): 180-191.

Bayne S (2015) What’s the matter with ‘technology-enhanced learning’? *Learning, Media And Technology*, 40(1), 5-20.

Bechhofer S et al. (2013) Why linked data is not enough for scientists. *Future Generation Computer Systems*, 29: 599-611.

Bennett S, Bishop A, Dalgarno B, Waycott J and Kennedy G. (2012) Implementing Web 2.0 technologies in higher education: a collective case study. *Computers and Education*, 59(2): 524-534.

Berardi, F. (2005) Info-Labour and Precarisation. Available at: <http://eicp.net/transversal/0704/bifo/en> (accessed 8 May 2017).

Berardi, F. (2009) *The Soul at Work: from Alienation to Autonomy*. Los Angeles, CA: MIT Press/Semiotexte.

Berners-Lee T, Hendler J, and Lassila O. (2001) The Semantic Web. *Scientific American*, 284(5): 34-43.

Bizer C, Heath T, and Berners-Lee T. (2009) Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*, 5(3): 1-22.

Borg E and Boyd Davis S. (2012) *The Thesis: Texts and Machines*. In R. Andrews, E. Borg, S. Boyd Davis, M. Domingo, and J. England (Eds.), *The SAGE Handbook of Digital Dissertations and Theses*. London: Sage.

Carmichael P et al. (2012) *Semantic Technologies for the Enhancement of Case-Based Learning*. ESRC End of Award Report, RES-139-25-0403-A. Swindon: ESRC

Carmichael P and Jordan K (2012) Semantic web technologies for education – time for a 'turn to practice'? *Technology, Pedagogy and Education*, 21(2): 153-169.

Carmichael P and Litherland K (2012) Transversality and Innovation: Prospects for Technology-Enhanced Learning in Times of Crisis. In Cole, D (ed), *Surviving Economic Crises through Education*. New York: Peter Lang, pp. 95-114.

Corneli J and Mikroyannidis A (2012) Crowdsourcing education on the Web: a rolebased analysis of online learning communities. In: Okada A, Connolly T and Scott P (eds) *Collaborative Learning 2.0: Open Educational Resources*. Hershey, PA: IGI Global, pp. 272–286.

Department for Education (2013) *National Curriculum in England: Computing Programmes of Study* London: Department for Education

De Roure D (2014) The future of scholarly communications. *Insights*, 27(3): 233-238.

De Waard A (2005) *Science Publishing And The Semantic Web, Or: Why Are You Reading This On Paper?* Proceedings of European Conference on the Semantic Web 2005, Heraklion, Greece.

Available at: <http://labs.elsevier.com/resources/adw/changingdoc/papers/deWaardECSW2005.pdf> (accessed 8 May 2017).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Dyer-Witheford N (2005) Cognitive capitalism and the contested campus. In: Cox G and Krysa. J (eds) *Engineering Culture: on the Author as (Digital) Producer* New York: Autonomedia, pp. 71-93.

Edwards R and Carmichael P (2012) Secret Codes: The Hidden Curriculum of Semantic Web Technologies. *Discourse*, 33(4): 575-590.

Edwards R, Tracy F and Jordan K (2011) Mobilities, moorings and boundary making in developing semantic technologies in educational practices. *Research in Learning Technology*, 19(3): 219-232.

European Commission (2016) Descriptors defining levels in the European Qualifications Framework (EQF). Available at: <https://ec.europa.eu/ploteus/content/descriptors-page> (accessed 8 May 2017).

Fox P and Hendler J (2009) Semantic e-Science: encoding meaning in next generation digitally enhanced science. In Hey T, Tansley S. and Tolle K(eds) *The fourth paradigm: data-intensive scientific discovery* Redmond, WA: Microsoft Research, pp. 147-153.

Frauenfelder M (2001) A Smarter Web. *MIT Technology Review*, Nov 1, 2001. Available at: <https://www.technologyreview.com/s/401241/a-smarter-web/> (accessed 8 May 2017).

Granieri G (2014). Culture now has two audiences: people and machines: A conversation with Ted Striphas. Available at: <https://medium.com/futurists-views/algorithmic-culture-culture-now-has-two-audiences-people-and-machines> (accessed 8 May 2017).

Healey M, Lannin L, Derounian J and Stibbe A (2012) *Rethinking final year projects and dissertations: Final report*. York: Higher Education Academy.

Heath T and Bizer C (2011) Linked Data: Evolving the Web into a Global Data Space. *Synthesis Lectures on the Semantic Web: Theory and Technology*, 1(1); 1-136.

Hey T, Tansley S and Tolle K (eds) (2009) *The Fourth Paradigm: Data-Intensive Scientific Research*. Redmond, WA: Microsoft Research.

Hoogerwef M, Loesch M, Schirrwagen J, Callaghan S, Manghi P, Iatropoulou K and Rettberg N (2013) Linking Data and Publications: Towards a Cross-Disciplinary Approach. *International Journal of Digital Curation*, 8(1): 244-254.

Huynh D, Karger D and Miller R. (2007) *Exhibit: lightweight structured data publishing*. Proceedings of WWW '07: Proceedings of the 16th International Conference on World Wide Web, Banff, Alberta, Canada New York, NY, USA.

Irwin A (1995) *Citizen Science: A Study of People, Expertise and Sustainable Development*. London: Routledge.

Jankowski N, Scharnhorst A, Tatum C, and Tatum Z (2013) Enhancing Scholarly Publications: Developing Hybrid Monographs in the Humanities and Social Sciences. *Scholarly and Research Communication*, 4(1). Available at: <http://src-online.ca/index.php/src/article/view/40/123> (accessed 8 May 2017).

Kitchin R (2016) Thinking critically about and researching algorithms. *Information, Communication and Society*, 20(1): 14-29.

Koper R (2004) Use of the Semantic Web to Solve Some Basic Problems in Education: Increase Flexible, Distributed Lifelong Learning, Decrease Teachers' Workload. *Journal of Interactive Media in Education*, 2004(6). Available at: <http://www-jime.open.ac.uk/articles/10.5334/2004-6-koper/> (accessed 8 May 2017).

Lytras M and Naeve A (2006) Semantic e-learning: synthesising fantasies. *British Journal of Educational Technology*, 37(3): 479-491.

Martinez-Garcia A, Morris S, Tscholl M, Tracy F and Carmichael P (2012) Case-Based Learning, Pedagogical Innovation, and Semantic Web Technologies. *IEEE Transactions on Learning Technologies*, 5(2): 104-116.

Marx, K (1973) *Grundrisse: Foundations of the Critique of Political Economy* (M. Nicolaus, Trans.). Harmondsworth: Penguin Books.

Mika P (2007a) *Social networks and the Semantic Web*. London: Springer.

Mika P (2007b) Ontologies are us: A unified model of social networks and semantics. *Journal of Web Semantics*, 5(1). Available at: <http://www.websemanticsjournal.org/index.php/ps/article/view/106> (accessed 8 May 2017).

Neyland D and Moellers N (2016) Algorithmic IF-THEN rules and the conditions and consequences of power. *Information, Communication and Society*, 20(1): 45-62.

Ozga J (2009) Governing Education through Data in England. *Journal of Education Policy*, 24(2):149-162.

Ozga J (2012). Governing Knowledge. *Globalisation, Societies and Education*, 10(4): 436-455.

Peters M (2011) Cognitive Capitalism and Educational Futures. In: Peters M and Bulut E (eds) *Cognitive Capitalism, Education and Digital Labour* New York: Peter Lang, pp. 245-258.

Porcello D and Hsi S (2013) Crowdsourcing and curating online education resources *Science* 341(6143): 240-241.

Ribes D and Bowker G (2008) Organizing for Multidisciplinary Collaboration: The Case of the Geosciences Network. In Olson G, Zimmerman A and Bos N (eds), *Scientific Collaboration on the Internet* Cambridge, MA.: MIT Press, pp. 311-330.

- Rimpiläinen S (2015) Multiple enactments of method, divergent hinterlands and production of multiple realities in educational research. *International Journal of Qualitative Studies in Education*, 28(2): 137-150.
- Selwyn N (2015) Data entry: Towards the critical study of digital data and education. *Learning, Media & Technology*, 40(1): 64-82.
- Shotton D, Portwin K, Klyne G, and Miles A (2009) Adventures in semantic publishing: exemplar semantic enhancement of a research article. *PLoS Computational Biology*, 5(4), e1000361. Available at: <https://doi.org/10.1371/journal.pcbi.1000361> (accessed 8 May 2017).
- Striphas T (2009) *The Late Age of Print: Everyday Book Culture from Consumerism to Control*. New York: Columbia University Press.
- Striphas T (2015) Algorithmic Culture. *European Journal of Cultural Studies*, 18(4-5): 395-412.
- Tracy F and Jordan K (2012) Students as designers of semantic web applications. *Technology, Pedagogy and Education*, 21(2): 171-188.
- Vratulis V and Dobson T (2008) Social negotiations in a wiki environment: A case study with preservice teachers. *Educational Media International*, 45(4): 285-294.
- Williamson B (2015a) Digital education governance: data visualization, predictive analytics, and real-time policy instruments. *Journal of Education Policy*, 31(2): 123-141.
- Williamson B (2015b) Governing Software: Networks, Databases and Algorithmic Power in the Digital Governance of Public Education. *Learning, Media & Technology*, 40(1): 83-105.
- Williamson B (2015c) Educating the smart city: Schooling smart citizens through computational urbanism. *Big Data & Society*, 2(2): 1-13.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Williamson B (2016) Political computational thinking: policy networks, digital governance and ‘learning to code’. *Critical Policy Studies*, 10(1): 39-58.

Williams R (1983) *Keywords: A Vocabulary of Culture and Society* (2nd ed.). Oxford: Oxford University Press.

For Peer Review

Funding

This work was supported by a grant from the UK ESRC/EPSRC Teaching Enhanced Learning Programme (<http://www.tlrp.org/tel/>) under research award RES-139-25-0403A.

Acknowledgements

The authors wish to acknowledge the contribution our colleagues and other participants who contributed to the ‘Ensemble’ project, particularly Professor Rob Walker (University of East Anglia), Dr. Robin Boast (formerly of the University of Cambridge Museum of Archaeology and Anthropology) and Ann Taylor (formerly of the University of Cambridge Department of Archaeology and Anthropology).

Author Biographies

Fran Tracy is Subject Head for Education and Early Childhood Studies at Liverpool John Moores University in the UK. Her research interests lie in the field of technology-enhanced learning in higher education and networked learning. She has a particular interest in working between disciplinary fields to support collaborative research. The majority of her research is participatory, involving undergraduate and postgraduate students as well as teaching staff in developing technologies alongside changes to curriculum and practice.

Patrick Carmichael is Professor of Education and Director of Research Development at the University of Bedfordshire in the UK and holds postgraduate degrees in the history and philosophy of science, and science education. His recent research activities and associated publications are concerned with technology-enhanced learning in higher education; participatory approaches to research, design and evaluation; and the development of ethical practice in emerging and trans-disciplinary settings. From 2008–2012 he was project director of ‘Ensemble’, a large research and development project exploring the educational potential of ‘linked data’ and the ‘semantic web’.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

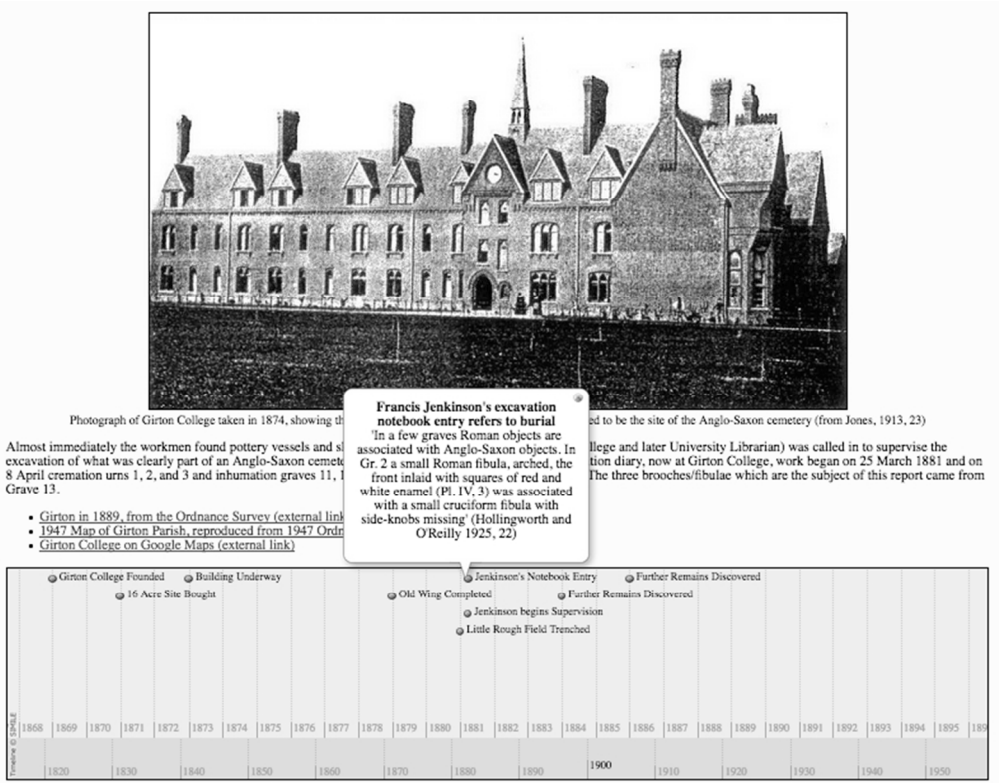


Figure 1: Part of the Data Aggregating Document including Interactive Timeline

Several early Anglo-Saxon cemeteries are now known in Cambridgeshire (see Map) - Edix Hill (Barrington A), Barrington B, Cambridge St John's, Gorton, Haslingfield, Linton Heath, Little Wilbraham, Oakington and Soham (Taylor, Duhig and Hines, 1997). Similarities have been noted, in particular with Oakington which "shares with a neighbouring cemetery at Gorton College a very high prevalence of cruciform and small long brooches in its female graves. [... These] brooch types are particularly characteristic of the costume of Anglian women in England, and their preponderance at Gorton and Oakington creates a marked contrast even with the cemeteries of South Cambridgeshire, where there is a considerably higher admixture of, for instance, saucer and disc brooches" (Taylor, Duhig and Hines 1997, 69). Many of these early cemeteries are also, like the Gorton cemetery, close to burial barrows, Roman roads, Roman villas, and parish boundaries (ibid., 70).

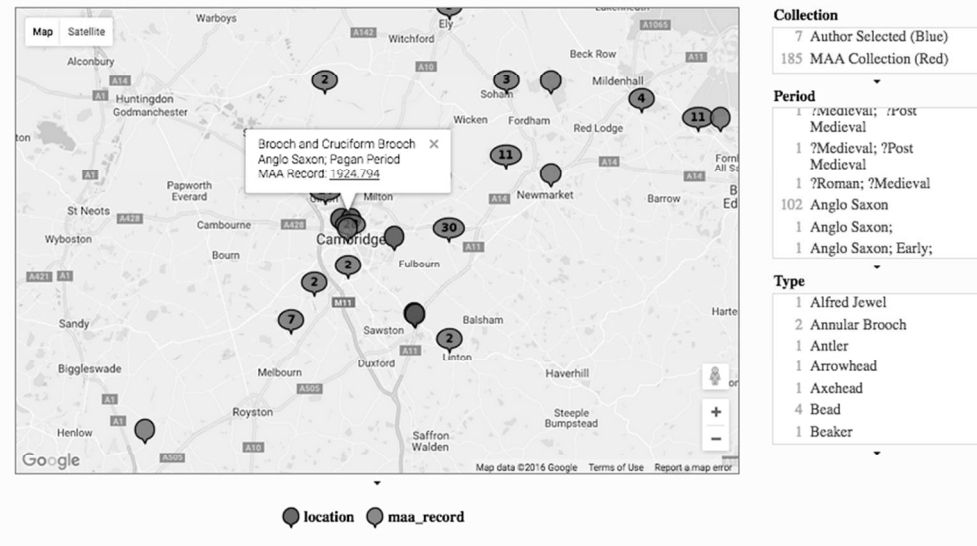


Figure 2: Part of Data Aggregating Document including interactive map with faceted search (right) and links to museum catalogue displayed for a selected site.

Table 1: Translating the Artefact Project into a Data Aggregating Document

Original Document	Data Aggregating Document	Potential Further Developments
Narrative text - contextual or interpretive	Narrative with hypertext links to external resources.	Citations to be made line links to online sources. More extensive 'semantic wikitext' implemented to allow machine readability of assertions in the narrative.
Narrative text - describing sequence of events	Interactive timeline annotated with key events.	Additional contextual information on timeline (e.g. related to developments in archaeology, key publications, other excavations and finds)
Images - single	Digitised versions of original images or copyright-free images from online sources.	Image search could be implemented to search for additional images of objects, sites, persons.
Images - collections	Searchable and sortable gallery of thumbnail images clickable to show full-size image and with links to museum catalogue entries.	Additional collections could be included if catalogues available and in appropriate formats.
Data tables	Tables sortable by column heading.	Hyperlinks to category systems, taxonomies or sources of data.
Maps - hand drawn and photocopied	Interactive 'Google' Maps with sites overlaid. Sites clickable to reveal museum catalogue entries. Full museum collection of Anglo-Saxon items now included.	Additional sources could be included; could allow switching between attributed age of artefacts and time of excavation.
Bibliography - manually compiled	Bibliography imported from BibTex file and exportable in variety of citation management formats.	Direct links to documents via DOI or ISBN numbers - less easily implemented with historical documents and archive content